



UvA-DARE (Digital Academic Repository)

Social inequality in shadow education: The role of high-stakes testing

Zwier, D.; Geven, S.; van de Werfhorst, H.G.

DOI

[10.1177/0020715220984500](https://doi.org/10.1177/0020715220984500)

Publication date

2020

Document Version

Final published version

Published in

International Journal of Comparative Sociology

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Zwier, D., Geven, S., & van de Werfhorst, H. G. (2020). Social inequality in shadow education: The role of high-stakes testing. *International Journal of Comparative Sociology*, 61(6), 412-440. <https://doi.org/10.1177/0020715220984500>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Social inequality in shadow education: The role of high-stakes testing

International Journal of
Comparative Sociology
2021, Vol. 61(6) 412–440
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0020715220984500
journals.sagepub.com/home/cos



Dieuwke Zwier , **Sara Geven**
and **Herman G. van de Werfhorst**

University of Amsterdam, The Netherlands

Abstract

Against the background of the worldwide expansion of shadow education, research shows that students from high socio-economic status (SES) backgrounds participate more in shadow education than students from disadvantaged SES backgrounds. We relate these social inequalities in shadow education participation to institutional features of educational systems. More specifically, we argue that the effect of socio-economic background on participation in shadow education will be stronger in countries characterized by high-stakes testing. Using data from the Programme for International Student Assessment for the year 2012 (PISA 2012), we show that higher SES students participate more in shadow education. For three out of four indicators of shadow education, this relationship is stronger in countries that are characterized by high-stakes testing but only when accounting for unobserved country differences.

Keywords

Cross-national research, high-stakes testing, PISA, shadow education, social inequalities

Introduction

The use of private, out-of-school educational activities has long been distinctive for East-Asian countries, such as Japan, Hong Kong, and South Korea, where the growth and monitoring of these educational activities took off after the Second World War (Hannum et al., 2019). More recently, participation in these educational activities has grown rapidly in countries all over the world—including low-income countries, such as Bangladesh, Cambodia, and Kenya; Eastern European and Central Asian countries; and North America and Western European countries (Bray, 2006; Silova, 2010). Private educational activities that occur outside formal education and that aim to enhance a student's educational achievement are commonly referred to as “shadow education” (Mori and Baker, 2010; Stevenson and Baker, 1992). Shadow education (SE) may take on different forms, including cram schools, one-to-one private tutoring, and examination preparatory courses. Despite cross-national variation regarding its scale, nature, and (historical) dynamics, SE is increasingly a worldwide phenomenon (Bray, 2006; Mori and Baker, 2010).

Corresponding author:

Dieuwke Zwier, Department of Sociology, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV, Amsterdam, The Netherlands.
Email: d.zwier@uva.nl

The expansion of SE may increase social inequalities in education (Bray, 2011, 2013; Buchmann et al., 2010). In various countries, students from high socio-economic status (SES) backgrounds participate more in SE than low-SES students (e.g. Buchmann et al., 2010; Davies, 2004; Park et al., 2016; Smyth, 2009; Stevenson and Baker, 1992). Common explanations are that higher SES parents are not only better equipped to pay for these supplementary learning activities, but also have more knowledge on how educational institutions work and are more intensively involved in their children's educational career (Buchmann et al., 2010; Park et al., 2016). Although prior research indicates that participation in SE has heterogeneous effects (Choi and Park, 2016) and its "effectiveness" depends on various factors—such as the quality of instruction, student motivation, and the national context (Bray, 2006)—SE generally improves a student's future educational opportunities and performance (Dang and Rogers, 2008). Consequently, social stratification in SE participation can be expected to raise educational inequalities.

Besides family background, institutional characteristics of educational systems are connected to the emergence and prevalence of SE. In this regard, "high-stakes testing" is a crucial institutional characteristic. High-stakes tests (HSTs) are standardized and centrally administered examinations with a decisive influence at critical educational transition points (Bray and Kwok, 2003; Hannum et al., 2019; Stevenson and Baker, 1992). The use of national standardized instruments for student certification, tracking, and/or graduation is increasing worldwide (EACEA/Eurydice, 2009; OECD, 2012). As HSTs are accompanied by intense competition for educational opportunities, and tight linkages between student achievement and future educational and occupational opportunities, scholars argue that HSTs fuel the prevalence of SE (Baker et al., 2001; Baker and LeTendre, 2005; Hannum et al., 2019; Stevenson and Baker, 1992).

Nonetheless, findings on the relationship between HSTs and SE are not conclusive. While prior cross-national research indicates that there is no relationship between HSTs and average national SE participation (Baker et al., 2001), single country studies posit HSTs as one of the main determinants producing greater demand for SE (e.g. Bray and Kwok, 2003; Lee et al., 2010; Zhang, 2014). So far, however, most research failed to examine how cross-national differences in institutional characteristics are related to the magnitude of social inequalities in the use of supplementary education in countries (see for an exception the contribution by Entrich, 2020 in this special issue). Consequently, what remains unclear is whether the national presence of HSTs similarly affects the use of SE activities by students from different socio-economic backgrounds.

Against this backdrop, we study the potential moderating influence of HSTs on the relationship between a student's socio-economic background and SE participation. More specifically, we expect that HSTs fuel the demand for SE among students from higher SES backgrounds. As higher SES parents may be more intensively involved in their child's educational career and better informed regarding the importance of HSTs (Grodsky, 2010; Park et al., 2011; Smyth, 2009), we argue that the (expected) positive relationship between parental SES and SE participation will be stronger in national educational systems characterized by high-stakes testing. To the best of our knowledge, together with Entrich (2020), we are the first to empirically study the moderating role of HSTs on the relationship between SES and SE participation.

We use data from the Programme for International Student Assessment (PISA) 2012 for 54 countries to study relationships between parental SES, HSTs, and participation in various types of supplementary education. Our results demonstrate that higher SES students are more prone to participate in SE. We find that high-stakes testing amplifies the relationship between parental SES and participation in different types of SE but only when all unobserved country differences are captured by country fixed effects.

Theoretical background

Socio-economic status and participation in shadow education

Both rational action theory and social-cultural reproduction theory offer explanations for variation in participation in (shadow) education across students from different socio-economic backgrounds (Smyth, 2009). Rational action theory views social class differences in educational decisions as outcomes of class-dependent calculations of the costs and benefits associated with continued participation in education. One influential rational action perspective—the relative risk aversion (RRA) mechanism—proposes that students try to avoid downward social class mobility, and this largely determines their educational strategies. That is, all students take their parents' social status as a reference for their own aspirations (Breen and Goldthorpe, 1997; Van De Werfhorst and Hofstede, 2007). As higher SES students have a higher reference level to ensure this goal of class maintenance, they are more likely to stay in the educational system than their low-SES counterparts with similar educational abilities. Class differences in the relative monetary costs of (shadow) education and in average academic performance levels—and, consequently, in expectations of educational success—further widen class differences in educational attainment (Breen and Goldthorpe, 1997). Although the costs of SE activities vary considerably across countries, SE participation is generally expensive. Lower SES families are often not able to invest in (good quality) SE (Bray, 2013; Smyth, 2009). In addition, lower SES families may be more likely to expect that their children will never get far enough in the educational system to benefit from initial investments at a later stage (Bray, 2006), making investments relatively more costly.

Socio-cultural reproduction theories focus on the unequal distribution of economic, social, and cultural resources across social classes, transmitted over generations (Bourdieu, 1977). With the process of educational expansion promoting social mobility (Breen, 2010), people from high social class backgrounds have to play a more active role in their children's educational careers to maintain social group boundaries. The use of SE may be one possible strategy for status maintenance among parents from high social class backgrounds (Lynch and Moran, 2006; Smyth, 2009).

More specifically, besides the financial ability to pay for it, participation in SE requires parents to gather information, to choose between various types and providers, and to monitor their children's progress (Park et al., 2011). Higher SES parents generally have more knowledge on the educational system, and, relatedly, the possibilities and potential benefits of supplementary learning activities (Buchmann et al., 2010). This makes them not only better equipped to make decisions regarding SE than low-SES parents, but they may also attach more value to this type of educational investments (Buchmann et al., 2010; Park et al., 2016; Smyth, 2009).

In this line of thought, SE investments can be seen as part of a wider strategy of “intensive parenting” or “concerted cultivation” (Lareau, 2003); a way of upbringing that is focused on the enhancement of children's educational success that is typically more found among families from high social class backgrounds (Buchmann et al., 2010; Byun et al., 2018; Park et al., 2016). Higher SES parents attach greater importance to academic performance, value a cognitive environment for their children, and closely monitor their children's activities, both in and outside school (Bray, 2013; Davies, 2004). They tend to structure their children's extracurricular activities (e.g. sports, arts, music) in educationally meaningful ways. This parenting style may extend to parental involvement in their children's supplementary education. In this way, the use of SE constitutes a form of cultural capital (Buchmann et al., 2010; Park et al., 2016).

While the rational action and the socio-cultural reproduction perspectives focus on different mechanisms for socio-economic inequalities in education, they both lead to the prediction that SES is positively related to students' SE participation. Prior empirical research finds support for this

relationship in various countries (e.g. in the United States (Buchmann et al., 2010), Canada (Davies, 2004), the United Kingdom (Ireson and Rushforth, 2005), Ireland (Smyth, 2009), Japan (Stevenson and Baker, 1992), South Korea (Kim and Lee, 2010), Vietnam (Dang, 2007), Hong Kong (Bray et al., 2014), China (Liu and Bray, 2016), Turkey (Tansel and Bircan, 2006), Georgia (Bregvadze, 2012), and various Eastern European and Central Asian countries (Silova, 2010)).

Hypothesis 1. Parental SES is positively associated with a student's participation in SE.

The role of high-stakes testing

Prior research links the presence of HSTs in the education system to the emergence and prevalence of SE (Baker et al., 2001; Bray, 2009, 2011; Buchmann et al., 2010; Hannum et al., 2019; Stevenson and Baker, 1992; Zhang, 2014). Scholars argue that HSTs enhance competitive behavior and the pressure to perform in education, and tighten the linkages between educational performance and labor market outcomes (Baker et al., 2001; Stevenson and Baker, 1992). An extreme form of these linkages is “diploma disease”: that is, employers relying heavily on degrees as an (indirect) signal of skills, causing educational certificates to have a large impact on hiring procedures. In this way, HSTs function as “public gatekeeper to education and labor market opportunities” (Baker et al., 2001: 3). Families are theorized to respond to the competitive pressures induced by the exam-oriented culture of high-stakes testing by seeking privilege for children outside of the formal education system, fueling the prevalence of SE (Hannum et al., 2019).

Empirical evidence on this hypothesized relationship between HSTs and SE at the national level is still mixed. One cross-national study, using cross-sectional data from the Trends in International Mathematics and Science Study (TIMSS) finds no relationship between high-stakes testing and the prevalence of SE among 12–14 years old children (Baker et al., 2001). However, national case studies suggest that high-stakes testing is one of the main mechanisms fueling the demand for private tutoring and other supplementary learning activities (Bray and Kwok, 2003; Lee et al., 2010; Zhang, 2014).

We propose that the competition that high-stakes tests induce may not work similarly for all students. Using insights from the socio-cultural reproduction framework, we argue that HSTs will especially bolster the use of SE among students from higher SES backgrounds. As outlined in the previous section, higher SES parents have more knowledge on the educational system (e.g. Smyth, 2009), and are actively involved in structuring their children's educational and extracurricular activities in order to ensure their academic success (Buchmann et al., 2010; Park et al., 2016, 2011). This may manifest itself, among other things, in more knowledge on the decisive influence of HSTs on students' future educational opportunities and life chances, and stronger beliefs about the presumed benefits of SE enrolment for boosting performances on HSTs (Buchmann et al., 2010).

Consequently, SE investments may be a more important strategy for parents from higher social class backgrounds to secure educational opportunities for their children (e.g. placement in highest tracks or elite schools) in a competitive educational environment in which test stakes are high than in an educational environment where test stakes are low. When test stakes are lower, incentives for high-SES parents to invest in SE for competitive advantage motives may be lower (Entrich, 2020). In contrast, parents from lower social class backgrounds may be less aware of the decisive role of high-stakes tests at educational transition points, and their SE investments are, therefore, expected to be less conditional on the presence of HSTs in the country.

Therefore, we expect differences in SE participation between students from socially advantaged and disadvantaged backgrounds to be larger in systems characterized by high-stakes testing, compared to systems without such testing.

Hypothesis 2. The positive relationship between SES and participation in SE will be stronger in national educational systems characterized by high-stakes testing.

Context

Despite the worldwide growth of SE, there are large cross-national variations in the prevalence and features of SE. First, the supply of SE varies across countries with respect to the following: (1) the class size—from personal tutoring to mass lectures, (2) the ages and qualifications of tutors—from formal teachers to students without formal training, and (3) the location and the use of (digital) technologies (Bray, 2006).

Second, countries differ in the relationship between SE and formal education. In some countries, such as Indonesia and Kazakhstan, semi-compulsory private tutoring is widespread; mainstream teachers pressure students to pay for out-of-school-time private lessons by educating only part of the formal curriculum in official school hours and the remaining part in supplementary lessons. This practice may be due to low wages of teachers in public education and/or relatively high levels of corruption in the country (Bray and Kwok, 2003; Silova, 2010; Zhang, 2014). Finally, participation in SE can fit either a “remedial” or “enrichment” user strategy, and the primary role that SE plays differs across countries. In most countries, SE has a “remedial” user strategy, implying that especially low-performing students participate in SE to eliminate achievement lags. In other countries, SE has an “enrichment” strategy, with a majority of above average performing participants; or is characterized by a combination of both strategies (Baker et al., 2001; Wolf, 2002). The enrichment strategy is primarily found in East-Asian countries such as South Korea (Byun et al., 2018; Choi and Park, 2016; Kim and Lee, 2010).

Data and methods

Data

PISA is an internationally standardized survey developed by the Organisation for Economic Co-operation and Development (OECD) that aims to measure educational performance in mathematics, reading (language), and science of 15-year-old students every 3 years. PISA 2012 focuses on mathematics performance, with reading and science as minor areas of assessment. We chose PISA 2012 as this wave covers a wider range of countries than previous rounds, and includes various items on SE, among which one indicator that allows us to identify fee-paying classes (see for more information, “Measures” section). Other PISA rounds do not include a measure that explicitly asks students about the usage of *paid* supplementary learning activities. Approximately, 510,000 students from 65 countries—including 34 OECD member countries and 31 partner countries—participated in PISA 2012 (OECD, 2014a). All countries, except for the Russian Federation, used a two-stage stratified sampling design. First, schools were sampled from a national sampling frame with probabilities proportional to the estimated number of 15-year-old students enrolled. Second, students were selected within schools with equal probability.

We exclude cases from the United Arab Emirates, Kazakhstan, Macau (China), Shanghai (China), Liechtenstein, Qatar, Perm (Russia), Serbia, and Tunisia due to missing internationally comparable information on country-level variables. Furthermore, we exclude Albania and Colombia due to the high shares of missing values on student-level variables.¹ The final sample includes 416,427 students from 54 countries. Depending on the proportion missing values on the dependent variable—primarily due to the survey’s rotated design—final sample sizes vary from $N=237,737$ to $N=260,634$ (54 countries).

We use sample weights to provide unbiased population estimates.² We normalize student weights at the country level, such that (1) the sum of weights across all countries is equal to the number of observations, (2) cases maintain the same proportional weights within each country, and (3) each country's contribution to the analyses is equal (OECD, 2009: 219).

Measures

Dependent variables. We measure *participation in SE* with four indicators from the PISA 2012 Student Context Questionnaires. The first two measures represent whether a student attends out-of-school-time lessons (OSL) in (1) language and (2) mathematics. Students provided the number of hours in OSL a week on a 5-point scale ranging from “no hours” to “6 or more hours a week.” Students also reported the number of hours they attend OSL in science and other subjects. We do not use these items, as science is often optional and the category “other subjects” is very broad, potentially causing (cross-national) variation in the question interpretation. The third SE indicator refers whether a student works with a personal tutor, paid or not. Finally, the fourth indicator measures whether a student attends out-of-school classes organized by a commercial company, paid for by his or her parents. These last two indicators are based on open-ended questions, where students indicated the number of hours a week they spend on these activities (range 0–30). For every measure, we recode the original responses into a dichotomous variable denoting whether a student participates at least some time in SE (1) or not (0). We also estimated models for the use intensity of SE (i.e. number of hours a week). These results are reported in Appendix 1 and are largely consistent with the main results.³

Note that only the fourth indicator makes the financial aspect of SE explicit—that is, “out of school classes organized by a commercial company, and *paid for by parents*” (OECD, 2013b: 233, emphasis added). These fee-paying classes are also referred to as the “high form” of SE (see Byun et al., 2018; Entrich, 2020). Although our other indicators (OSL and private tutoring) may also partly capture public learning activities outside formal education, we believe it is important to also analyze these indicators, as children may not always know whether/that their out-of-school-time classes are paid for. Hence, if we would solely focus on measures that explicitly capture “high forms” of SE, we may underestimate possible social inequalities in SE. Moreover, SE may at times be implicitly paid for. For example, there might be (private) schools who offer OSL or personal tutors, yet who also request higher tuition fees. All in all, it is likely that participation in OSL and private tutoring is also stratified (see Byun et al., 2018). Hence, we conduct separate analyses for the different SE indicators. In this way, we can indicate potential differences between broad indicators (OSL and personal tutoring) and narrow indicator (commercial company classes). In addition, we use prior research to interpret the results in a more comprehensive way.

The PISA 2012 Student Context Questionnaires have a “rotated design.” This implies that the SE measures are only asked to two-third of the students, and one-third of the observations within each country is missing by design (N/A) (OECD, 2014b: 58–61). For all countries, we examine the share of observations *not* missing by design (invalid/missing). The share of invalid/missing values on OSL mathematics and OSL language generally does not exceed 4 percent, except for eight countries that have 4–6 percent missing/invalid observations. Compared to the first two indicators, the share of missing/invalid values on personal tutor and commercial company lessons is generally higher, especially in some South-American and Asian countries (e.g. Argentina, Brazil, Peru, Indonesia, Jordan, and Turkey; $\geq 10\%$ of all observations). However, we do not impute missing values on the dependent variable as this can add noise to the estimates (Von Hippel, 2007).

Independent variables. The independent variable of primary interest at the student level is *socio-economic status*. We measure SES with the “PISA index of Economic, Social and Cultural status”

(ESCS). This verified index combines indicators for parental occupation, parental education, and home possessions. The latter includes items on family wealth possessions, cultural possessions, home educational resources, and the number of books at home. The ESCS-index is standardized such that the mean of OECD member countries equals 0 and the SD equals 1 (OECD, 2014b).⁴

To operationalize *high-stakes testing*, we focus on central examinations with Bishop's (1997) definition of "Curriculum-based External Exit Examinations" (CBEEEs) (cf. Baker et al., 2001; Entrich, 2020). The use of nationally standardized exams with a decisive influence at educational transitions is an important indicator of a culture of high-stakes testing. Bishop distinguishes five characteristics of CBEEEs: (1) the exam produces signals of student achievement that have real consequences for students, (2) student achievement is defined relative to an external standard, (3) the exam is organized by discipline, (4) it signals multiple levels of achievement in the subject (not pass/fail), and (5) it is administered to almost all secondary school students (Bishop, 1997).

We derive this indicator from previous research (Bishop, 1997; Bol et al., 2014; Fuchs and Wößmann, 2007; Wößmann, 2003; Wößmann et al., 2009) and supplement and update it with other data sources (EACEA/Eurydice, 2009, 2015; EP-Nuffic, 2015; OECD, 2008, 2012, 2013c; UNESCO-IBE, 2012). The vast majority of countries scores either 0 (no central exams) or 1 (central exams) on this variable. However, in four countries—Australia, Canada, Germany, and the United States—there are no nationally centralized examinations. These countries score 0.81, 0.51, 0.44 and 0.09, respectively, representing the proportion of subnational regions where these examinations are present (see Bol et al., 2014; Wößmann et al., 2009, for a similar approach). Table 1 summarizes country-level scores on CBEEEs.

Control variables. At the student level, we control for a number of socio-demographic factors and educational performance that potentially confound the relationship between parental SES and student participation in SE (e.g. Byun et al., 2018; Park et al., 2016).

More specifically, we control for the student's *age*, as the age of PISA's target population varies between 15 years and 3 months and 16 years and 2 months at the start of the testing period (OECD, 2014b: 460). Students' ages systematically vary across countries, and students' age may also impact their educational performance, and therefore, their SE participation. We also account for the student's *migration background*. We use the PISA index of immigrant background and distinguish between (1) native citizens (students who have at least one parent born in the country of assessment), (2) first-generation immigrants (students born outside the country of assessment whose parents were also born in another country), and (3) second-generation students (students born in the country of assessment whose parent(s) were born in another country; OECD, 2014b: 307). We also include the student's *gender* (female = 1) and *grade* (years of formal education).

In addition, we control for students' educational performance, as performance is related to a student's social background as well as to his or her motivation to make use of SE activities (remediation vs enrichment). We measure *student's educational performance* with plausible values (PVs). During PISA's computer-based assessments, students answer a subset of test items on mathematics, reading (test language) and science, rather than all items. Instead of obtaining a point estimate to measure the student's educational performance levels, a posterior probability distribution based on the student's item responses is estimated. PVs are randomly drawn values from this distribution. Hence, PVs are a representation of "the range of abilities that a student might reasonably have, given the student's item responses" (Wu, 2005: 115). For each educational domain—language, mathematics, and science—PISA 2012 includes five PVs (OECD, 2009; Wu, 2005). Following the *PISA Data Analysis Manual* (OECD, 2009), we compute our statistical models separately for each PV, and use Rubin's combination rules to average these estimates, and to obtain the final mean estimates and corresponding standard errors (OECD, 2014b; Wu, 2005). As PVs are ranging from circa 0 to 900, we divided them by 100 to ease interpretation of the coefficients.

Table 1. Descriptive statistics country-level high-stakes testing and tracking age.

Country abbreviation	Country	Central exams	Tracking age
ARG	Argentina	0	15
AUS	Australia	0.81	16
AUT	Austria	0	10
BEL	Belgium (Flemish and French)	0	12
BGR	Bulgaria	1	13
BRA	Brazil	0	15
CAN	Canada	0.51	16
CHE	Switzerland	0	12
CHL	Chile	0	16
CRI	Colombia	1	15
CZE	Czech Republic	1	11
DEU	Germany	0.44	10
DNK	Denmark	1	16
ESP	Spain	0	16
EST	Estonia	1	15
FIN	Finland	1	16
FRA	France	1	15
GBR	The United Kingdom	1	16
GRC	Greece	0	15
HKG	Hong Kong	1	15
HRV	Croatia	1	14
HUN	Hungary	1	11
IDN	Indonesia	1	15
IRL	Ireland	1	15
ISL	Iceland	0	16
ISR	Israel	1	15
ITA	Italy	1	14
JOR	Jordan	1	16
JPN	Japan	1	15
KOR	Korea	1	14
LTU	Lithuania	1	16
LUX	Luxembourg	1	13
LVA	Latvia	1	16
MEX	Mexico	0	15
MNE	Republic of Montenegro	0	15
MYS	Malaysia	1	15
NLD	The Netherlands	1	12
NOR	Norway	1	16
NZL	New Zealand	1	16
PER	Peru	0	16
POL	Poland	1	16
PRT	Portugal	0	15
ROU	Romania	1	14
RUS	Russian Federation	1	16
SGP	Singapore	1	12
SVK	Slovak Republic	1	11

(Continued)

Table 1. (Continued)

Country abbreviation	Country	Central exams	Tracking age
SVN	Slovenia	1	14
SWE	Sweden	0	16
TAP	Chinese Taipei	1	15
THA	Thailand	1	15
TUR	Turkey	1	11
URY	Uruguay	0	11
USA	The United States	0.09	16
VNM	Vietnam	1	15

Sources: Bishop (1997), Bol et al. (2014), EACEA/Eurydice (2009, 2015), EP-Nuffic (2015), Fuchs and Wößmann (2007), OECD (2008, 2012, 2013a, 2013c), UNESCO-IBE (2012), Wößmann (2003), Wößmann et al. (2009).

At the country level, we control for *tracking age*. National educational systems differ in multiple ways, and these institutional features relate to educational inequalities by socio-economic background. Two main features that relate to such inequalities are (1) the level of national standardization (among which the standardization of output, by means of central exams) and (2) the level of differentiation (Van De Werfhorst and Mijs, 2010). The latter often refers to between-school tracking—that is, the selection and allocation of students into separate ability tracks or streams (e.g. vocational and academic) (LeTendre et al., 2003).⁵ In this respect, prior research demonstrated that educational equalities by SES are larger in highly differentiated educational system, versus comprehensive systems, especially when tracking occurs at a younger age (Bol et al., 2014; Brunello and Checchi, 2007; Hanushek and Wößmann, 2006; Van De Werfhorst and Mijs, 2010). We measure the level of between-school tracking in a country by the first age of selection (*tracking age*) (cf. Hanushek and Wößmann, 2006). We obtain this information from the OECD (2013a; see Table 1). Tracking age ranges in our data between 10 years and 16 years. In the statistical models, the variable is mean-centered to ease interpretation of the results.

Furthermore, we control for a cross-level interaction between *tracking age* and a student's socio-economic background. Similar to central exams, between-school tracking may especially fuel the demand for SE among high-SES students (see also Enrich, 2020). Students' track allocation strongly determines students' future educational pathways and tracking thus raises the stakes of educational transitions. Especially, higher SES parents may have the financial resources and cultural knowledge to “navigate” these educational institutions. Consequently, they are more likely to invest in SE around critical transition points in order to improve their child(ren)'s performance, and to influence track allocation. In contrast, countries with lower levels of differentiation may offer less opportunities to gain competitive advantages through supplementary educational activities in allocation processes (Enrich, 2020). Hence, we want to make sure that the interaction between central exams and a student's SES holds after accounting for the interaction between tracking age and a student's SES.

Summary statistics for all student-level variables are presented in Table 2. Only 3.5 percent of the sample has (a) missing value(s) on one or more of the student-level predictors. We exclude these cases from analysis.

Methods

PISA 2012 data have a hierarchical structure, with students i nested in schools j , and schools nested in countries k . Three-level random effects linear probability models (LPMs) are used to correct

Table 2. Descriptive statistics individual-level variables.

	Mean	Min.	Max.	SD	N
Participation in shadow education					
Out-of-school-time lessons language (yes = 1)	0.37	0	1		266,507
Out-of-school-time lessons mathematics (yes = 1)	0.48	0	1		266,852
Personal tutoring (yes = 1)	0.34	0	1		245,985
Commercial company classes (yes = 1)	0.32	0	1		242,864
Socio-economic status (ESCS-index)	-0.61	-5.95	3.69	1.27	410,140
Educational performance					
PV mathematics 1	4.58	0.20	9.25	1.03	416,427
PV mathematics 2	4.58	0.44	9.32	1.03	416,427
PV mathematics 3	4.58	0.44	9.12	1.03	416,427
PV mathematics 4	4.58	0.25	9.12	1.03	416,427
PV mathematics 5	4.58	0.51	8.97	1.03	416,427
PV reading 1	4.66	0.03	9.05	1.02	416,427
PV reading 2	4.66	0.01	8.81	1.02	416,427
PV reading 3	4.66	0.01	8.84	1.02	416,427
PV reading 4	4.66	0.04	8.81	1.02	416,427
PV reading 5	4.66	0.05	9.02	1.02	416,427
Gender (female = 1)	0.50	0	1	0.50	416,427
Age	15.80	15.17	16.33	0.29	416,312
Grade	9.76	7	13	0.75	415,554
Immigration background					
Native citizen (ref.)	0.93	0	1		405,059
First-generation immigrant	0.03	0	1		405,059
Second-generation immigrant	0.05	0	1		405,059

Source: PISA 2012 own calculations.

SD: standard deviation; PV: plausible value.

Sample weights are used. The descriptive statistics are calculated over the total sample $N=416,427$ (54 countries), as sample sizes vary between the models due to missing values on the dependent variable.

standard errors for this clustering (Hox, 2010; Rabe-Hesketh and Skrondal, 2005). For each of the four SE indicators, we estimate similar models.⁶

First, we report the intercept only model (Model 0), and calculate school-level and country-level intraclass correlations (ICC). Model 1 presents the analysis of the effects of SES (Hypothesis 1), the student-level sociodemographic control variables, and central exams on participation in SE (0=no, 1=yes). Models 2–5 include the cross-level interaction between SES and central exams (Hypothesis 2). The model specification for Model 2 is as follows:

$$\begin{aligned}
 P(Y_{ijk}=1) = & \beta_0 + \beta_{1-5} X_{ijk} + \beta_6 \text{SES}_{ijk} + \beta_7 \text{central exams}_k + \\
 & \beta_8 \text{SES}_{ijk} \times \text{central exams}_k + u_{0k} + \\
 & v_{0jk} + u_{6k} \text{SES}_{ijk} + e_{ijk}
 \end{aligned}$$

With Y_{ijk} referring to participation in a particular type of SE, X_{ijk} to a vector of student-level socio-demographic control variables (gender, age, grade, immigration background), and u_{0k} and v_{0jk} to the country-level and school-level variance, respectively. The model includes a random slope of SES at the country level ($u_{6k} \text{SES}_{ijk}$), to allow the effect of SES to differ across countries,

and to avoid downward biased standard errors when estimating cross-level interactions (Heisig and Schaeffer, 2019).

We build on Model 2 in several ways. In Model 3, we add the PVs to assess whether the findings hold when controlling for performance. In the models predicting participation in OSL mathematics, personal tutoring, and commercial company lessons, we include five PVs measuring the student's performance in mathematics; in the models predicting OSL language, we use PVs measuring performance in reading (language). While it is important to control for the student's educational performance, we acknowledge that performance may be endogenous on SE participation. As we use cross-sectional data, it is impossible to include a measure of performance *prior* to SE participation. Consequently, observed performance can be both a driver and a consequence of SE participation. For this reason, we add performance separately to the model, and abstain from drawing strong conclusions about the effect of student performance on SE participation based on our analysis.

In Model 4, we assess whether including the age at which students are tracked alters the results. We add both the main effect of tracking age and a cross-level interaction between SES and tracking age. Next to the level of between-school tracking, countries vary in other (un)observed ways that may alter the relationship between SES and SE participation. It is difficult to measure all potentially relevant factors for the relationship between SES, central exams, and supplementary education with country-level indicators. Therefore, we use a fixed effects approach in Model 5 to assess the cross-level interaction between SES and central exams. Rather than estimating a country-level random intercept, we estimate a two-level LPM with students i clustered in schools j and include country fixed effects to control for all between-country unobserved heterogeneity (see for a similar approach, Bol et al., 2014). This model is specified as follows:

$$P(Y_{ijk}=1)=\beta_0+\beta_{1-5}X_{ij}+\beta_6SES_{ij}+\beta_7SES_{ij}\times\text{central exams}_k+\sum_k^{k-1}\beta_xC_k+v_{0j}+e_{ij}$$

In this model specification, all country variance is captured by the country fixed effects (denoted by $\sum_k^{k-1}\beta_xC_k$), so we cannot estimate the main effect of central exams. Our primary focus, however, is to assess whether the relationship between SES and SE participation is stronger in systems with central exams (indicative of high-stakes testing). We assess this potential variation in the strength of the effect of our individual-level variable of interest (SES) by including a cross-level interaction between SES and central exams. This interaction effect can be estimated in a model with country fixed effects, as it varies between individuals within the same country.

Results

Descriptive results

Figure 1 summarizes national descriptive statistics of the proportion of students participating in different forms of SE, and the average number of hours a week that students spend on these activities. The figures show large cross-country differences in participation in SE activities, in particular, personal tutoring and commercial company lessons. We also find differences between forms of SE: in most countries, national participation in OSL mathematics is high compared to participation in other SE activities. We want to emphasize that we can make no distinction between OSL lessons provided by public schools and those provided by private institutions.

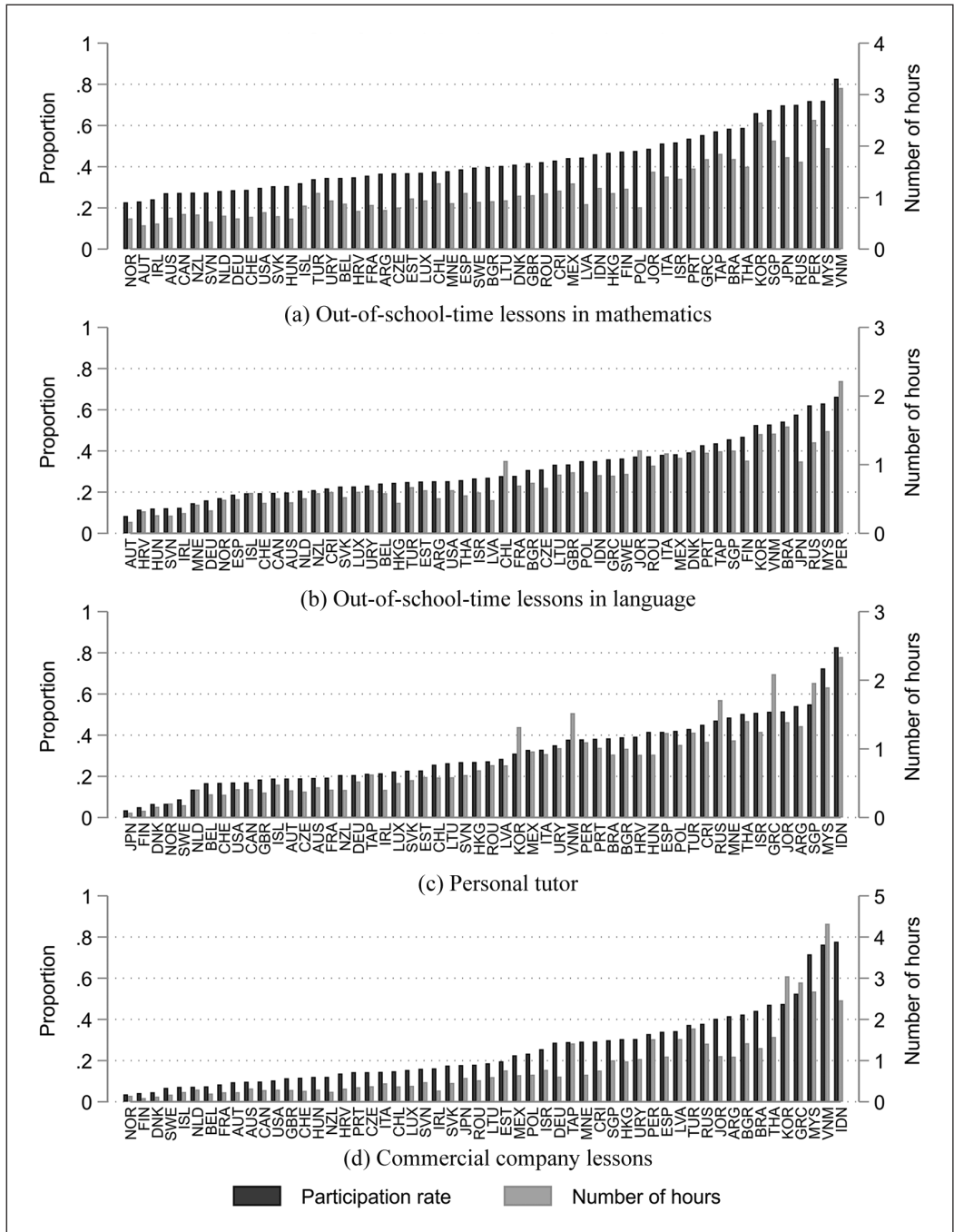


Figure 1. National proportions of students participating in shadow education (left axis), and the number of hours a week students on average spend on these lessons (right axis), countries in ascending order. (a) Out-of-school-time lessons in mathematics. (b) Out-of-school-time lessons in language. (c) Personal tutor. (d) Commercial company lessons.

Source: PISA 2012 own calculations.

Sampling weights are used. (a) $N = 260,634$; (b) $N = 260,267$; (c) $N = 240,729$; (d) $N = 240,729$.

There appears to be a correlation between different SE measures at the country level. In countries where students' enrolment in OSL is relatively low (e.g. Austria, the Czech Republic, or Norway) or high (e.g. Vietnam, Peru, or Tunisia), participation rates in personal tutoring or commercial company classes are also low or high, respectively. However, not every country takes a similar position on the different rankings. This may be due to cross-national variation in the supply of SE and/or the interpretation of PISA questions (Bray and Kobakhidze, 2014). For instance, Japanese participation rates in OSL are far above average but this country has the lowest participation rate in personal tutoring. While the use of SE is widespread in Japan, one-to-one tutoring is less common than large cram schools (e.g. Bray, 2006; Stevenson and Baker, 1992). In addition, Denmark has one of the lowest participation rates in personal tutoring and commercial lessons, while taking a high centrist position on the country rankings of OSL in mathematics/language. Possibly, OSL in Denmark include supplementary lessons that the public education system offers free of charge. Finally, the high prevalence of personal tutoring and commercial lessons in Indonesia is slightly counter-intuitive. Despite the fact that quite a lot of Indonesian students tend to participate in supplementary lessons supplied by their own teachers after regular school hours (Kim and Lee, 2010), Indonesia is not known for a very high prevalence of SE. These patterns support the importance of conducting separate analyses for the different indicators of SE.⁷

Multilevel regression results

Tables 3 to 6 show the results of the multilevel regression models for the different SE indicators. Model 0 indicates that the proportion of variance at the school level varies between 5.6 percent (personal tutoring) and 6.6 percent (OSL language). The ICCs at the country level are higher: between 8.3 percent (OSL mathematics) and 15.0 percent (commercial company lessons) of the variance in the dependent variable is at the country level.⁸

In Model 1, we estimate the coefficients of SES, central exams (i.e. indicator of high-stakes testing), and the socio-demographic student-level variables. First, we find a positive association between parental SES (ESCS-index) and three out of four measures of SE. In line with our expectation (Hypothesis 1), high-SES students are more likely to participate in SE activities than low-SES students. For every one-unit (i.e. one *SD*) increase on the (mean-standardized) ESCS-index, predicted chances to participate in OSL mathematics, personal tutoring and commercial company lessons on a weekly basis increase by 3.1, 4.2, and 3.2 percentage-points, respectively. The relationship between SES and OSL language is not significantly different from zero. Although these social background effects on SE participation do not appear to be very large at first glance, our estimates can be interpreted as relatively conservative, given the inability to distinguish between paid and unpaid supplementary education for OSL and personal tutoring.

This model also shows that the use of central exams is not significantly associated with any of the four indicators of SE. Consistent with prior cross-national comparative work (Baker et al., 2001; Entrich, 2020), this implies that the general demand for SE does not seem to be larger in countries characterized by high-stakes testing, compared to countries not using such testing. Moreover, this result is consistent across the various SE measures, from broad indicators potentially partly capturing public supplementary learning activities provided by schools to "high forms" of SE (i.e. commercial company lessons), explicitly capturing the financial costs of participation.

Regarding the sociodemographic control variables, both first- and second-generation immigrants tend to have a higher probability to participate in SE than native students. Furthermore, female students have a higher probability of participating in OSL mathematics, personal tutoring and commercial company lessons, compared to male students. We find a negative association between a student's grade and SE participation, though this association loses statistical

Table 3. Results multilevel regression models out-of-school-time lessons in mathematics.

	(0)	(1)	(2)	(3)	(4)	(5)
Fixed part						
Female		0.018** (0.006)	0.018** (0.006)	0.012* (0.006)	0.018** (0.006)	0.018** (0.006)
Age of student		0.006 (0.008)	0.006 (0.008)	0.003 (0.007)	0.006 (0.008)	0.007 (0.009)
Grade		-0.038*** (0.011)	-0.038*** (0.011)	-0.019 (0.014)	-0.038*** (0.011)	-0.038*** (0.005)
Immigrant background (ref. = native)						
First-generation		0.056** (0.018)	0.056** (0.018)	0.049** (0.019)	0.056** (0.018)	0.055*** (0.016)
Second-generation		0.046*** (0.011)	0.046*** (0.011)	0.042*** (0.012)	0.046*** (0.011)	0.046*** (0.012)
SES		0.031*** (0.004)	0.019* (0.008)	0.030*** (0.008)	0.019* (0.008)	0.013*** (0.004)
Central exams		0.065 (0.044)	0.065 (0.044)	0.081* (0.041)	0.064 (0.041)	
Central exams × SES			0.018+ (0.009)	0.019* (0.009)	0.018+ (0.009)	0.025*** (0.006)
Performance (PVMATH)				-0.051*** (0.015)		
Tracking age					0.021* (0.009)	
Tracking age × SES					-0.003+ (0.002)	
Country FE	No	No	No	No	No	Yes
Constant	0.429*** (0.019)	0.645** (0.203)	0.645** (0.203)	0.754*** (0.202)	0.646** (0.202)	0.611*** (0.140)
Random part						
$\hat{\sigma}_{u0}^2$ (country-level)	0.020*** (0.004)	0.024*** (0.005)	0.024*** (0.005)	0.023*** (0.005)	0.023*** (0.005)	
$\hat{\sigma}_{v0}^2$ (school-level)	0.014*** (0.003)	0.013*** (0.002)	0.013*** (0.002)	0.013*** (0.003)	0.013*** (0.002)	0.013*** (0.001)
$\hat{\sigma}_e^2$ (student-level)	0.208*** (0.005)	0.207*** (0.005)	0.207*** (0.005)	0.206*** (0.005)	0.207*** (0.005)	0.207*** (0.001)
$\hat{\sigma}_{u6}^2$ (SES)		0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	
BIC	4,980,071.9	4,952,465.6	4,952,474.3		4,952,494.1	4,959,339.3
Log likelihood	-2,490,011.0	-2,476,158.0	-2,476,156.1		-2,476,153.5	-2,479,276.8

Source: PISA 2012 own calculations.

SES: socio-economic status, FE: fixed effects, BIC: Bayesian information criterion.

Standard errors in parentheses. Normalized sampling weights are used (robust standard errors reported).

Tracking age is mean-centered. N=260,634 (15,965 schools, 54 countries).

+p < 0.10; *p < 0.05; **p < 0.05; ***p < 0.001.

Table 4. Results multilevel regression models out-of-school-time lessons in language.

	(0)	(1)	(2)	(3)	(4)	(5)
Fixed part						
Female		0.001 (0.008)	0.001 (0.008)	-0.006 (0.008)	0.001 (0.008)	0.002 (0.005)
Age of student		0.005 (0.008)	0.005 (0.008)	0.001 (0.008)	0.005 (0.008)	0.006 (0.009)
Grade		-0.039*** (0.012)	-0.039*** (0.012)	-0.018 (0.012)	-0.039*** (0.012)	-0.040*** (0.005)
Immigrant background (ref. = native)						
First-generation		0.084*** (0.018)	0.084*** (0.018)	0.076*** (0.019)	0.084*** (0.018)	0.082*** (0.017)
Second-generation		0.027* (0.011)	0.027* (0.011)	0.024** (0.008)	0.027* (0.011)	0.025+ (0.013)
SES		-0.001 (0.003)	-0.007 (0.006)	0.006 (0.007)	-0.007 (0.006)	-0.004 (0.004)
Central exams		0.037 (0.045)	0.037 (0.045)	0.056 (0.042)	0.036 (0.041)	
Central exams × SES			0.009 (0.008)	0.010 (0.008)	0.009 (0.008)	0.008 (0.006)
Performance (PVREAD)				-0.061*** (0.013)		
Tracking age					0.025*** (0.008)	
Tracking age × SES					-0.003+ (0.001)	
Country FE	No	No	No	No	No	Yes
Constant	0.316*** (0.019)	0.580** (0.215)	0.580** (0.215)	0.720** (0.242)	0.582** (0.213)	0.541*** (0.138)
Random part						
$\hat{\sigma}_{u0}^2$ (country-level)	0.020*** (0.004)	0.020*** (0.004)	0.020*** (0.004)	0.020*** (0.003)	0.018*** (0.003)	
$\hat{\sigma}_{v0}^2$ (school-level)	0.015*** (0.003)	0.013*** (0.003)	0.013*** (0.003)	0.013*** (0.003)	0.013*** (0.003)	0.014*** (0.001)
$\hat{\sigma}_e^2$ (student-level)	0.194*** (0.005)	0.194*** (0.005)	0.194*** (0.005)	0.192*** (0.006)	0.194*** (0.005)	0.194*** (0.002)
$\hat{\sigma}_{u6}^2$ (SES)		0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	
BIC	4,725,935.1	4,708,058.3	4,708,069.4		4,708,086.3	4,714,375.9
Log likelihood	-2,362,942.6	-2,353,954.3	-2,353,953.6		-2,353,949.6	-2,356,795.2

Source: PISA 2012 own calculations.

SES: socio-economic status, FE: fixed effects, BIC: Bayesian information criterion.

Standard errors in parentheses. Normalized sampling weights are used (robust standard errors reported).

Tracking age is mean-centered. $N=260,267$ (15,973 schools, 54 countries).

+ $p < 0.10$; * $p < 0.05$; ** $p < 0.05$; *** $p < 0.001$.

Table 5. Results multilevel regression models personal tutor.

	(0)	(1)	(2)	(3)	(4)	(5)
Fixed part						
Female		0.052*** (0.010)	0.052*** (0.010)	0.041*** (0.012)	0.052*** (0.010)	0.053*** (0.005)
Age of student		0.019*** (0.006)	0.019*** (0.006)	0.013* (0.006)	0.019*** (0.006)	0.020* (0.008)
Grade		-0.059*** (0.009)	-0.059*** (0.009)	-0.025** (0.009)	-0.059*** (0.009)	-0.059*** (0.005)
Immigrant background (ref. = native)						
First-generation		0.054*** (0.012)	0.054*** (0.012)	0.043*** (0.011)	0.054*** (0.012)	0.054*** (0.016)
Second-generation		0.028* (0.014)	0.028* (0.014)	0.024 (0.015)	0.028* (0.014)	0.026* (0.012)
SES		0.042*** (0.005)	0.033*** (0.007)	0.054*** (0.007)	0.032*** (0.007)	0.016*** (0.004)
Central exams		0.016 (0.050)	0.016 (0.050)	0.045 (0.043)	0.015 (0.049)	
Central exams × SES			0.013 (0.010)	0.016 (0.010)	0.013 (0.010)	0.035*** (0.006)
Performance (PVMATH)				-0.097*** (0.011)		
Tracking age					0.004 (0.011)	
Tracking age × SES					-0.004 (0.002)	
Country FE	No	No	No	No	No	Yes
Constant	0.308*** (0.023)	0.545*** (0.135)	0.545*** (0.135)	0.761*** (0.165)	0.545*** (0.135)	0.773*** (0.128)
Random part						
$\hat{\sigma}_{u0}^2$ (country-level)	0.027*** (0.006)	0.031*** (0.006)	0.031*** (0.006)	0.024*** (0.005)	0.030*** (0.006)	
$\hat{\sigma}_{v0}^2$ (school-level)	0.012*** (0.003)	0.010*** (0.002)	0.010*** (0.002)	0.010*** (0.002)	0.010*** (0.002)	0.011*** (0.001)
$\hat{\sigma}_e^2$ (student-level)	0.174*** (0.014)	0.171*** (0.013)	0.171*** (0.013)	0.166*** (0.013)	0.171*** (0.013)	0.172*** (0.002)
$\hat{\sigma}_{u6}^2$ (SES)		0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	
BIC	3,761,069.1	3,684,500.2	3,684,511.0		3,684,533.8	3,712,777.6
Log likelihood	-1,880,509.7	-1,842,175.7	-1,842,175.0		-1,842,174.0	-1,855,998.5

Source: PISA 2012 own calculations.

SES: socio-economic status, FE: fixed effects, BIC: Bayesian information criterion.

Standard errors in parentheses. Normalized sampling weights are used (robust standard errors reported).

Tracking age is mean-centered. N=240,729 (15,914 schools, 54 countries).

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.001$.

Table 6. Results multilevel regression models commercial company lessons.

	(0)	(1)	(2)	(3)	(4)	(5)
Fixed part						
Female		0.024*** (0.007)	0.024*** (0.007)	0.014+ (0.008)	0.024*** (0.007)	0.024*** (0.005)
Age of student		0.026*** (0.005)	0.026*** (0.005)	0.021*** (0.005)	0.026*** (0.005)	0.026** (0.009)
Grade		-0.063*** (0.003)	-0.063*** (0.003)	-0.032*** (0.005)	-0.063*** (0.003)	-0.062*** (0.005)
Immigrant background (ref. = native)						
First-generation		0.057*** (0.011)	0.057*** (0.011)	0.047*** (0.009)	0.057*** (0.011)	0.059*** (0.016)
Second-generation		0.041*** (0.007)	0.041*** (0.007)	0.036*** (0.008)	0.041*** (0.007)	0.042*** (0.012)
SES		0.032*** (0.004)	0.028*** (0.006)	0.046*** (0.007)	0.027*** (0.006)	0.020*** (0.004)
Central exams		0.032 (0.054)	0.032 (0.054)	0.057 (0.049)	0.031 (0.054)	
Central exams × SES			0.007 (0.008)	0.009 (0.008)	0.007 (0.008)	0.012+ (0.006)
Performance (PVMATH)				-0.085*** (0.011)		
Tracking age					0.009 (0.011)	
Tracking age × SES					-0.002 (0.002)	
Country FE	No	No	No	No	No	Yes
Constant	0.248*** (0.024)	0.417*** (0.077)	0.417*** (0.077)	0.600*** (0.105)	0.417*** (0.076)	0.606*** (0.135)
Random part						
$\hat{\sigma}_{u0}^2$ (country-level)	0.030*** (0.008)	0.037*** (0.009)	0.037*** (0.009)	0.033*** (0.009)	0.037*** (0.009)	
$\hat{\sigma}_{v0}^2$ (school-level)	0.012*** (0.003)	0.011*** (0.002)	0.011*** (0.002)	0.010*** (0.002)	0.011*** (0.002)	0.011*** (0.001)
$\hat{\sigma}_e^2$ (student-level)	0.160*** (0.014)	0.158*** (0.014)	0.158*** (0.014)	0.155*** (0.014)	0.158*** (0.014)	0.158*** (0.002)
$\hat{\sigma}_{u6}^2$ (SES)		0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	
BIC	3,416,367.8	3,372,677.4	3,372,689.2		3,372,712.7	3,385,218.5
Log likelihood	-1,708,159.1	-1,686,264.4	-1,686,264.1		-1,686,263.5	-1,692,219.3

Source: PISA 2012 own calculations.

SES: socio-economic status, FE: fixed effects, BIC: Bayesian information criterion.

Standard errors in parentheses. Normalized sampling weights are used (robust standard errors reported). Tracking age is mean-centered. $N=237,737$ (15,899 schools, 54 countries).

+ $p < 0.10$; * $p < 0.05$; ** $p < 0.05$; *** $p < 0.001$.

significance for OSL mathematics and language when including educational performance (see Model 3). Finally, there is a positive relationship between a student's age and participation in personal tutoring and commercial company lessons.

While there does not appear to be a general association between central exams and SE enrolment, high-states testing may still bolster SE participation among high-SES students. We test Hypothesis 2 by including a cross-level interaction between SES and central exams in Models 2–5. The results for Model 2 show that the positive relationship between SES and OSL mathematics is stronger in countries with central exams ($0.019 + 0.018 \times \text{central exams}$), compared to countries without central exams (see Table 3, Model 2). Note, however, that this effect is only borderline statistically significant ($p=0.06$). In the models for OSL language, personal tutoring, and commercial company lessons, the coefficients for the $\text{SES} \times \text{central exams}$ interaction term are in the expected direction but do not reach statistical significance (see Tables 4 to 6). So far, the results in Model 2 show no convincing evidence for our expectation that the positive relationship between SES and participation in shadow expectation is stronger in HSTs-characterized countries (Hypothesis 2). Note, however, that some relevant factors (e.g. student performance, tracking age, and other country differences) are not accounted for in this model specification yet.

In Model 3, we include educational performance to Model 2. We generally find a negative relationship between educational performance and participation in various SE activities. With every one unit (=100 points on PVs) increase in educational performance, predicted chances to participate in SE decrease with 5.1 (OSL mathematics), 6.1 (OSL language), 9.7 (personal tutor), and 8.5 (commercial company lessons) percentage-point. After the inclusion of educational performance, the positive association between SES and the different measures of SE becomes somewhat stronger. This is possibly due to the positive correlation between SES and educational performance ($r=0.48, p < 0.01$), and negative association between educational performance and SE, causing the relationship between SES and SE to be suppressed when performance is not included in the model. The results for the cross-level interactions between SES and central exams are similar to those in Model 2. Again, the interaction term $\text{SES} \times \text{central exams}$ is significantly positive in the models of OSL mathematics ($b=0.019, p < 0.05$). Furthermore, we find no statistically significant cross-level interaction between SES and HSTs for OSL language, personal tutoring, and commercial company lessons.

Model 4 includes tracking age (i.e. the first age of selection). The results indicate that the higher the first age of selection, the higher predicted chances of participation in OSL mathematics and language. This relationship seems to be slightly weaker among students from higher socio-economic backgrounds, though interaction terms between tracking age \times SES are very close to zero and statistically significant at $p=0.10$. More importantly, including tracking age and its cross-level interaction with SES do not alter the results for the moderating role of central exams in the models for OSL in mathematics. Tracking age is not significantly associated with personal tutoring and commercial company lessons.

Finally, Model 5 includes country fixed effects. When statistically controlling for all between-country variance, cross-level interaction terms between SES and central exams are positive and statistically significant for three out of four measures of SE. The positive relationship between SES and SE participation is stronger in countries with central exams for OSL mathematics ($0.013 + 0.025 \times \text{central exams}$), personal tutoring ($0.016 + 0.035 \times \text{central exams}$), and commercial company lessons ($0.020 + 0.012 \times \text{central exams}$). Note, however, that the cross-level interaction between SES and central exams for commercial company lessons is only borderline statistically significant ($p=0.06$). For OSL language, both the estimated main effect of SES and the interaction between SES and central exams do not significantly differ from zero.⁹

Figure 2 summarizes the results of the country fixed effects models (Model 5) by showing average marginal effects (AMEs) of SES on participation in different forms of SE in countries

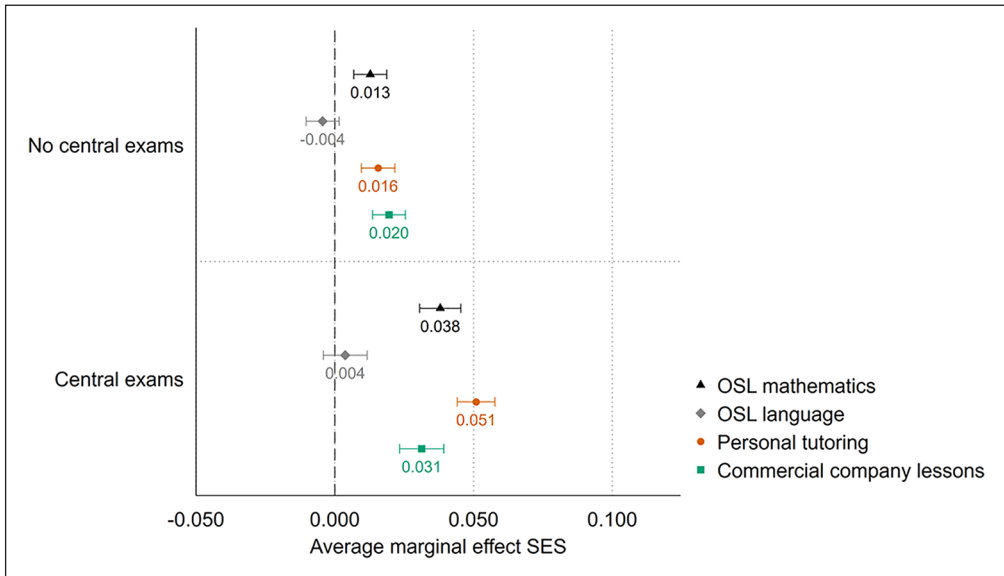


Figure 2. Average marginal effects SES on participation in different types of shadow education in countries with and without central exams.

Source: PISA 2012 own calculations.

90% confidence intervals displayed. The average marginal effects displayed are based on Model 5 (see Tables 3 to 6).

with and without central exams. In comparison with the results presented in Models 2–4, the results in the country fixed effects models provide more support for Hypothesis 2. The figure illustrates that the positive effect of SES on the predicted probability of SE enrolment is stronger in countries characterized by central exams, compared to countries without central exams, for OSL mathematics, personal tutoring, and commercial company lessons. This suggests that social inequalities in different types of SE are more pronounced in countries characterized by central exams. Unobserved cross-country differences may have suppressed the moderating role of central exams for some types of supplementary learning (i.e. commercial company, private tutoring) in the previous models.

Conclusion and discussion

The worldwide expansion of SE may have far-reaching implications for educational systems in terms of social inequalities. Prior research in various countries demonstrated that high-SES students are more prone to participate in SE (e.g. Buchmann et al., 2010; Davies, 2004; Kim and Lee, 2010; Silova, 2010). An institutional characteristic that is often connected to the emergence and prevalence of SE is high-stakes testing (Baker and LeTendre, 2005; Buchmann et al., 2010; Stevenson and Baker, 1992). However, so far, no studies have examined whether HSTs moderate the relationship between students' socio-economic background and SE participation. Using data from PISA 2012, we studied the extent to which students from higher SES backgrounds participate more in SE, and whether this relationship is strengthened by the presence of HSTs.

Confirming previous studies (e.g. Buchmann et al., 2010; Entrich, 2020; Smyth, 2009; Stevenson and Baker, 1992), our results showed that higher SES students generally participate more in various SE activities. However, also note the findings by Entrich (2020) in this special issue, demonstrating

that in some countries no social inequality or even a reversed SES gap in SE enrolment exists. Social inequalities in SE may translate into inequalities in educational performance and attainment. While the strength and direction of the association between SE enrolment and educational outcomes depends on students' background characteristics, the type and use intensity of SE, and the national context, SE participation generally improves a student's future educational outcomes (Choi and Park, 2016; Dang and Rogers, 2008; Smyth, 2009).

In line with other comparative work (Baker et al., 2001; Enrich, 2020), the presence of central exams (indicative of a culture of high-stakes testing) was not related to students' SE participation in general. However, and in line with our expectation, we found evidence—when accounting for unobserved country differences—that in countries characterized by high-stakes testing, social inequalities are more pronounced in OSL in mathematics, personal tutoring, and commercial company lessons.

In a related study using PISA 2012 data in this special issue, Enrich (2020) found no evidence for higher social inequalities in commercial company lessons and personal tutoring in countries characterized by high-stakes testing. While these findings may seem at odds with our findings, similar to Enrich (2020), we also found no support for the moderating effect of HSTs on the relationship between SES and commercial company lessons or personal tutoring in models in which we did not account for unobserved country differences (i.e. models without country fixed effects). Possibly unobserved economic, cultural, and/or institutional characteristics of countries are related to both the absence (or presence) of HSTs in a country, as well as the presence (or absence) of socio-economic inequalities in SE. Not accounting for these unobserved country differences may then suppress the moderating role of HSTs on socio-economic inequalities in SE. Note that these factors are likely to be different from those included in the study by Enrich (2020) (e.g. world region, level of income inequality, education enrolment rates, and national institutionalization of SE). Future research should further examine how high-stakes testing—and the accompanying incentives to gain comparative advantages through SE participation—relate to the magnitude of social inequalities in the use of supplementary education.

Our study has some limitations. As mentioned earlier, some of our SE measures are broad, and do not distinguish paid from unpaid supplementary education. PISA questions about out-of-school-time classes and personal tutoring are modified almost every cycle. The latest PISA assessments (2015 and 2018) include (different) questions on participation in various types of “additional instruction” in the “Education career questionnaire.” Unfortunately, only part of the countries participated in these optional surveys. Moreover, these rounds do not make the fee-paying nature of SE explicit. The worldwide expansion of SE makes it increasingly relevant for international surveys on education such as PISA to further develop and specify questions about SE. Although researchers will always have to account for cross-national variations in the characteristics of SE and potential country differences in the interpretation of survey items, this will enhance the possibilities for cross-national research on SE.

Another data-related issue is that the PISA sample is limited to 15-year-old students, while HSTs usually take place at the end of secondary education when students are approximately 17–19 years old (OECD, 2012). Hence, for students in the PISA sample—with either a lower or higher SES—the presence of high-stakes tests may not *yet* affect their SE participation, as they may not have to take these tests in the near future. If feasible, future research should focus on a slightly older sample of students, and develop a more fine-grained operationalization of high-stakes testing (i.e. number of years to the test¹⁰).

Furthermore, due to the cross-sectionality of our data, we cannot draw any causal conclusions based on our findings. Ideally, we should have used a longitudinal design to examine our research question. However, there were insufficient changes in the extent to which countries use central

exams in the period for which cross-nationally comparative data on HSTs and SE is available. Moreover, items on SE have changed almost every PISA-cycle, impeding the longitudinal analyses of SE with PISA data.

Relatedly, we controlled for educational performance in some of our models. While recognizing that performance may be endogenous on SE participation, performance is also known to be an important determinant of SE enrolment. Hence, leaving this confounder out of the regression equation can result in omitted variable bias. We find that performance is negatively associated with participation in SE activities, suggesting that SE has a remedial purpose for most students. Note, however, that the relationship between performance and SE can differ between countries. In this regard, prior research showed that in some countries, especially in Eastern Asia, the enrichment strategy (i.e. high-performing students are more likely to use SE than low-performing students) is dominant (Baker et al., 2001; Byun et al., 2018). Furthermore, the magnitude of social inequalities in SE is found to be related to incentives for higher-performing students to use SE (Enrich, 2020), which may be higher in competitive systems characterized by high-stakes testing.

If longitudinal information on performance and SE is available, an interesting avenue for future research would therefore be to further examine the relationship between SES, performance, HSTs, and SE participation. It may, for instance, be that social disparities in SE are especially pronounced among high-performing students: when test stakes are high, higher SES parents may invest in SE, independent of their children's performance. In contrast, investments in SE of low-SES parents may be more conditional on performance; they may only invest in SE when their children's performance is low, not when it is relatively high.

Despite these limitations, this study has, again, called attention to the social inequalities in participation in SE. Furthermore, we have contributed to comparative work on the role of institutional characteristics—in specific, high-stakes testing—on these social inequalities.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was made possible through a Vici-grant awarded to Herman G. van de Werfhorst by the Netherlands' Organisation for Scientific Research (NWO), grant number 453-14-017 and a Veni-grant awarded to Sara Geven by the Netherlands' Organisation for Scientific Research (NWO), grant number 016.Veni.195.125.

ORCID iD

Dieuwke Zwier  <https://orcid.org/0000-0002-5898-5557>

Notes

1. For Albania, all information on socio-economic status (SES) is missing, and for Colombia, the share of missing values on shadow education measures exceeds 60 percent.
2. These sample weights correct for (1) the over- or undersampling of some population strata for national reporting purposes, (2) lack of accuracy or no updated size measures on the school sampling frame, and (3) weight adjustment for school and student non-response (OECD, 2009).
3. For this purpose, we recoded the categorical items for OSL in mathematics and language into (pseudo-)interval variables using the midpoint of each category. For instance, "2 or more but less than 4 hours a week" becomes "3." The last category "6 or more hours a week" is open-ended; due to the right skewed distribution of the items, we assign the value "7" to this category. Personal tutoring and commercial company lessons are top-coded to 10 to preclude the small share of respondents (0.9 and 1.9 percent, respectively) that filled in an unrealistically high number of hours a week to affect the results. All results are reported in Appendix 1, Tables 8 to 11, and are largely consistent with the main results. The most

notable difference with our main results (Tables 3 to 6) is that the positive cross-level interaction between SES and central exams also reaches statistical significance in the models for OSL language (see Table 9 in Appendix, Models 2–5).

4. As a robustness check, we examined which component of the ESCS-index—parental occupation, parental education, or home possessions (wealth)—is most important in predicting participation in shadow education. Overall, the home possessions summary index is most strongly and consistently associated with all shadow education indicators.
5. Note that we refer to between-school tracking. Tracking occurs not only between but also within schools (see Chmielewski, 2014; LeTendre et al., 2003). However, within-school tracking is often less institutionalized at the country level, and countries that generally hold stronger cultural beliefs about the acceptability of formally allocating students to different tracks at an early age (LeTendre et al., 2003).
6. We prefer LPMs over binary logistic models as linear probability estimates offer a more intuitive interpretation than (log-)odds, are easier to compare across models, and LPMs require less computational power and time than logistic models, especially when using multiple imputation techniques and including random slopes. We repeated the main analyses using logistic regression models, and the substantive results were consistent with the results presented here.
7. Moreover, we re-estimated specific models 55 times, while dropping one of the 55 countries each time, to assure that our main results do not depend on a single country. In addition, the DFBETA diagnostic is used to assess the influence of potential country-level outliers on the estimated cross-level interaction between SES and high-stakes tests (HSTs). DFBETAs are calculated as the difference between the estimated regression coefficient in the full sample and in the sample without the specific country, divided by the standard error of the estimate in the sample excluding the country. Influential units are countries that exceed the critical value of $|2/\sqrt{N_j}|$, with N_j referring to the number of countries (i.e. 54). In Models 2, three to four countries exceed this critical threshold for the different types of SE. When re-estimating this model excluding these countries, the substantive results remain similar. In the model for OSL mathematics (four influential countries), the positive relationship between SES and OSL mathematics is stronger in countries with central exams ($0.009 + 0.025 \times \text{central exams}$; $p < 0.001$). One notable difference with our main results for Model 2 is that the cross-level interaction between SES and central exams also reaches statistical significance in the model for personal tutoring ($b = 0.018$; $p < 0.05$), after excluding three influential countries.
8. Intraclass correlations (ICCs) are not reported in Tables 3 to 6. The ICC at school level is defined as $\frac{\hat{\sigma}_{v0}^2}{\hat{\sigma}_{u0}^2 + \hat{\sigma}_{v0}^2 + \hat{\sigma}_e^2}$ and the ICC at country level as $\frac{\hat{\sigma}_{u0}^2}{\hat{\sigma}_{u0}^2 + \hat{\sigma}_{v0}^2 + \hat{\sigma}_e^2}$, following the first method discussed by Hox (2010: 33–34).
9. In the models for SE use intensity (i.e. hours a week, see Note 3), we also find a slightly positive cross-level interaction between SES and central exams in the models for OSL language (see Table 9 in Appendix). In Model 5, the main effect of SES approximates zero closely, and is slightly negative in countries without HSTs (-0.021), versus slightly positive in countries with HSTs ($-0.021 + 0.038 = 0.017$).
10. Note that this may differ across students within a country, as the moment of a high-stakes test may depend on a student's track-level in a tracked education system.

References

- Baker DP, Akiba M, LeTendre GK, et al. (2001) Worldwide shadow education: Outside-school learning, institutional quality of schooling, and cross-national mathematics achievement. *Educational Evaluation and Policy Analysis* 23(1): 1–17.
- Baker DP and LeTendre GK (2005) *National Differences, Global Similarities: World Culture and the Future of Schooling*, 1st edn. Stanford, CA: Stanford Social Sciences.

- Bishop JH (1997) The effect of national standards and curriculum-based exams on achievement. *The American Economic Review* 87(2): 260–264.
- Bol T, Witschge J, Van de Werfhorst HG, et al. (2014) Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 countries. *Social Forces* 92(4): 1545–1572.
- Bourdieu P (1977) Cultural reproduction and social reproduction. In: Karabel J and Halsey AH (eds) *Power and Ideology in Education*. New York: Oxford University Press, pp. 487–511.
- Bray M (2006) Private supplementary tutoring: Comparative perspectives on patterns and implications. *Compare: A Journal of Comparative and International Education* 36(4): 515–530.
- Bray M (2009) Confronting the shadow education system: What government policies for what private tutoring? Available at: <http://unesdoc.unesco.org/images/0018/001851/185106e.pdf>
- Bray M (2011) *The Challenge of Shadow Education: Private Tutoring and Its Implications for Policy Makers in the European Union*. Luxembourg: European Commission.
- Bray M (2013) Shadow education: Comparative perspectives on the expansion and implications of private supplementary tutoring. *Procedia—Social and Behavioral Sciences* 77(2013): 412–420.
- Bray M and Kobakhidze MN (2014) Measurement issues in research on shadow education: Challenges and pitfalls encountered in TIMSS and PISA. *Comparative Education Review* 58(4): 590–620.
- Bray M and Kwok P (2003) Demand for private supplementary tutoring: Conceptual considerations, and socio-economic patterns in Hong Kong. *Economics of Education Review* 22(6): 611–620.
- Bray M, Zhan S, Lykins C, et al. (2014) Differentiated demand for private supplementary tutoring: Patterns and implications in Hong Kong secondary education. *Economics of Education Review* 38(2014): 24–37.
- Breen R (2010) Educational expansion and social mobility in the 20th century. *Social Forces* 89(2): 365–388.
- Breen R and Goldthorpe JH (1997) Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society* 9(3): 275–305.
- Bregvadze T (2012) Analysing the shadows: Private tutoring as a descriptor of the education system in Georgia. *International Education Studies* 5(6): 80–89.
- Brunello G and Checchi D (2007) Does school tracking affect equality of opportunity? New international evidence. *Economic Policy* 22(52): 781–861.
- Buchmann C, Condron DJ and Roscigno VJ (2010) Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces* 89(2): 435–462.
- Byun S, Chung HJ and Baker DP (2018) Global patterns of the use of shadow education: Student, family, and national influences. *Research in the Sociology of Education* 20: 71–105.
- Chmielewski AK (2014) An international comparison of achievement inequality in within- and between-school tracking systems. *American Journal of Education* 120(3): 293–324.
- Choi Y and Park H (2016) Shadow education and educational inequality in South Korea: Examining effect heterogeneity of shadow education on middle school seniors' achievement test scores. *Research in Social Stratification and Mobility* 44(2016): 22–32.
- Dang HA (2007) The determinants and impact of private tutoring classes in Vietnam. *Economics of Education Review* 26(6): 683–698.
- Dang HA and Rogers FH (2008) The growing phenomenon of private tutoring: Does it deepen human capital, widen inequalities, or waste resources? *The World Bank Research Observer* 23(2): 161–200.
- Davies S (2004) School choice by default? Understanding the demand for private tutoring in Canada. *American Journal of Education* 110(3): 233–255.
- EACEA/Eurydice (2009) National testing of pupils in Europe: Objectives, organisation and use of results. Available at: <https://op.europa.eu/en/publication-detail/-/publication/df628df4-4e5b-4014-adbd-2ed54a274fd9/language-en>
- EACEA/Eurydice (2015) Languages in secondary education: An overview of national tests in Europe. Available at: <https://op.europa.eu/en/publication-detail/-/publication/62ac43c3-dac4-11e5-8fea-01aa75ed71a1/language-en>
- Entrich S (2020) Worldwide shadow education and social inequality: Explaining differences in the socio-economic gap in access to shadow education across 63 societies. *International Journal of Comparative Sociology*.

- EP-Nuffic (2015) Foreign education systems. Available at: <https://www.nuffic.nl/>
- Fuchs T and Wößmann L (2007) What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics* 32(2–3): 433–464.
- Grodsky E (2010) Learning in the shadows and in the light of day: A commentary on shadow education, American style: test preparation, the SAT and college enrollment. *Social Forces* 89(2): 475–481.
- Hannum E, Ishida H, Park H, et al. (2019) Education in East Asian societies: Postwar expansion and the evolution of inequality. *Annual Review of Sociology* 45: 625–647.
- Hanushek EA and Wößmann L (2006) Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal* 116(510): 63–77.
- Heisig JP and Schaeffer M (2019) Why you should always include a random slope for the lower-level variable involved in a cross-level interaction. *European Sociological Review* 35(2): 258–279.
- Hox J (2010) *Multilevel Analysis: Techniques and Applications*, 2nd edn. Hove: Routledge.
- Ireson J and Rushforth K (2005) Mapping and evaluating shadow education. Available at: <http://image.guardian.co.uk/sys-files/Education/documents/2005/04/07/tutoring.pdf>
- Kim S and Lee J-H (2010) Private tutoring and demand for education in South Korea. *Economic Development and Cultural Change* 58(2): 259–296.
- Lareau A (2003) *Unequal Childhoods: Class, Race, and Family Life*. Berkeley, CA: University of California Press.
- Lee CJ, Lee H and Jang HM (2010) The history of policy responses to shadow education in South Korea: Implications for the next cycle of policy responses. *Asia Pacific Education Review* 11(1): 97–108.
- LeTendre GK, Hofer BK and Shimizu H (2003) What is tracking? Cultural expectations in the United States, Germany, and Japan. *American Educational Research Association* 40(1): 43–89.
- Liu J and Bray M (2016) Determinants of demand for private supplementary tutoring in China: Findings from a national survey. *Education Economics* 25(2): 205–218.
- Lynch K and Moran M (2006) Markets, schools and the convertibility of economic capital: The complex dynamics of class choice. *British Journal of Sociology of Education* 27(2): 221–235.
- Mori I and Baker D (2010) The origin of universal shadow education: What the supplemental education phenomenon tells us about the postmodern institution of education. *Asia Pacific Education Review* 11(1): 36–48.
- OECD (2008) How do OECD countries take stock of progress and performance in education systems? Evidence and issues (No. 33). Available at: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC\(2008\)33&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC(2008)33&docLanguage=En)
- OECD (2009) PISA data analysis manual: SPSS, second edition. Available at: https://www.oecd-ilibrary.org/education/pisa-data-analysis-manual-spss-second-edition_9789264056275-en
- OECD (2012) Education at a Glance 2012: OECD indicators. Available at: https://www.oecd-ilibrary.org/education/education-at-a-glance-2012_eag-2012-en
- OECD (2013a) 2. Selecting and grouping students. In: *PISA 2012 results: What makes schools successful? Resources, policies, and practices*: Vol. IV, pp. 155–181. Available at: https://www.oecd-ilibrary.org/education/pisa-2015-results-volume-ii/selecting-and-grouping-students_9789264267510-9-en
- OECD (2013b) PISA 2012 Assessment and Analytical Framework: Mathematics, reading, science, problem solving and financial literacy. *OECD Report*. Available at: https://www.oecd-ilibrary.org/education/pisa-2012-assessment-and-analytical-framework_9789264190511-en
- OECD (2013c) Synergies for better learning: An international perspective on evaluation and assessment. Available at: https://www.oecd-ilibrary.org/education/synergies-for-better-learning-an-international-perspective-on-evaluation-and-assessment_9789264190658-en
- OECD (2014a) *PISA 2012 Results in Focus: What 15-Year-Olds Know and What They Can Do with What They Know*, Revised edn. Available at: <https://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-I.pdf>
- OECD (2014b) PISA 2012 technical report. Available at: <http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>
- Park H, Buchmann C, Choi J, et al. (2016) Learning beyond the school walls: Trends and implications. *Annual Review of Sociology* 42(1): 231–252.

- Park H, Byun S and Kim K (2011) Parental involvement and students' cognitive outcomes in Korea: Focusing on private tutoring. *Sociology of Education* 84(1): 3–22.
- Rabe-Hesketh S and Skrondal A (2005) *Multilevel and Longitudinal Modeling Using Stata*. Berkeley, CA: Stata Press.
- Silova I (2010) Private tutoring in Eastern Europe and Central Asia: Policy choices and implications. *Compare: A Journal of Comparative and International Education* 40(3): 327–344.
- Smyth E (2009) Oxford review of education buying your way into college? Private tuition and the transition to higher education in Ireland Buying your way into college? Private tuition and the transition to higher education in Ireland. *Oxford Review of Education* 35(1): 1–22.
- Stevenson DL and Baker DP (1992) Shadow education and allocation in formal schooling: Transition to University in Japan. *American Journal of Sociology* 97(6): 1639–1657.
- Tansel A and Bircan F (2006) Demand for education in Turkey: A Tobit analysis of private tutoring expenditures. *Economics of Education Review* 25(3): 303–313.
- UNESCO-IBE (2012) World data on education: Seventh edition 2010-11. Available at: <http://www.ibe.unesco.org/en/document/world-data-education-seventh-edition-2010-11>
- Van de Werfhorst HG and Hofstede S (2007) Cultural capital or relative risk aversion? Two mechanisms for educational inequality compared. *The British Journal of Sociology* 58(3): 391–415.
- Van de Werfhorst HG and Mijs JJB (2010) Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology* 36: 407–428.
- Von Hippel PT (2007) Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology* 37(1): 83–117.
- Wößmann L (2003) Central exit exams and student achievement: International evidence. In: Peterson PE and West MR (eds) *No Child Left Behind? The Politics and Practice of School Accountability*. Washington, DC: Brookings Institution Press, pp. 292–323.
- Wößmann L, Luedemann E, Schuetz G, et al. (2009) *School Accountability, Autonomy and Choice around the World*. Cheltenham: Edward Elgar.
- Wolf RM (2002) Extra-school instruction in mathematics and science. In: Robitaille DF and Beaton AE (eds) *Secondary Analysis of the TIMSS Data*, 1st edn. Boston, MA: Kluwer, pp. 331–341.
- Wu M (2005) The role of plausible values in large-scale surveys. *Studies in Educational Evaluation* 31(2–3): 114–128.
- Zhang W (2014) The demand for shadow education in China: Mainstream teachers and power relations. *Asia Pacific Journal of Education* 34(4): 436–454.

Appendix I

Supplementary analyses

Table A1. Results multilevel regression models out-of-school-time lessons in mathematics (use intensity: hours a week).

	(0)	(1)	(2)	(3)	(4)	(5)
Fixed part						
Female		0.026 (0.018)	0.026 (0.018)	0.017 (0.021)	0.026 (0.018)	0.027 (0.020)
Age of student		0.028 (0.028)	0.028 (0.028)	0.024 (0.027)	0.028 (0.028)	0.031 (0.034)
Grade		-0.090** (0.031)	-0.090** (0.031)	-0.062 (0.045)	-0.090** (0.031)	-0.090*** (0.019)
Immigrant background (ref. = native)						
First-generation		0.138* (0.064)	0.138* (0.064)	0.128* (0.064)	0.138* (0.064)	0.137* (0.056)
Second-generation		0.095*** (0.029)	0.095*** (0.029)	0.090*** (0.026)	0.095*** (0.029)	0.096* (0.046)
SES		0.106*** (0.019)	0.042+ (0.024)	0.059* (0.026)	0.042+ (0.024)	0.032* (0.013)
Central exams		0.136 (0.180)	0.136 (0.180)	0.160 (0.173)	0.131 (0.171)	
Central exams × SES			0.094* (0.036)	0.096** (0.036)	0.094** (0.036)	0.118*** (0.022)
Performance (PVMATH)				-0.076 (0.059)		
Tracking age					0.070* (0.036)	
Tracking age × SES					-0.007 (0.007)	
Country FE	No	No	No	No	No	Yes
Constant	1.143*** (0.074)	1.496* (0.691)	1.496* (0.691)	1.659* (0.671)	1.499* (0.686)	1.153* (0.523)
Random part						
$\hat{\sigma}_{u0}^2$ (country-level)	0.291*** (0.073)	0.404** (0.130)	0.404** (0.130)	0.399** (0.131)	0.388** (0.125)	
$\hat{\sigma}_{v0}^2$ (school-level)	0.189*** (0.042)	0.164*** (0.037)	0.164*** (0.037)	0.170*** (0.039)	0.164*** (0.037)	0.179*** (0.014)
$\hat{\sigma}_{\epsilon}^2$ (student-level)	2.820*** (0.159)	2.808*** (0.158)	2.808*** (0.158)	2.803*** (0.159)	2.808*** (0.158)	2.811*** (0.040)
$\hat{\sigma}_{u\epsilon}^2$ (SES)		0.020*** (0.005)	0.018*** (0.004)	0.018*** (0.004)	0.018*** (0.004)	
BIC	14,880,163.9	14,850,287.2	14,850,294.6		14,850,316.9	14,863,325.6
Log likelihood	-7,440,057.0	-7,425,068.8	-7,425,066.3		-7,425,064.9	-7,431,269.9

Source: PISA 2012 own calculations.

SES: socio-economic status, FE: fixed effects, BIC: Bayesian information criterion.

Standard errors in parentheses. Normalized sampling weights are used (robust standard errors reported).

Tracking age is mean-centered. N=260,634 (15,965 schools, 54 countries).

+p < 0.10; *p < 0.05; **p < 0.05; ***p < 0.001.

Table A2. Results multilevel regression models out-of-school-time lessons in language (use intensity: hours a week).

	(0)	(1)	(2)	(3)	(4)	(5)
Fixed part						
Female		0.002 (0.015)	0.002 (0.015)	0.060** (0.023)	0.002 (0.015)	0.003 (0.018)
Age of student		0.014 (0.019)	0.014 (0.019)	-0.003 (0.020)	0.014 (0.019)	0.017 (0.031)
Grade		-0.109** (0.035)	-0.109** (0.035)	-0.038 (0.047)	-0.109** (0.035)	-0.110*** (0.017)
Immigrant background (ref. = native)						
First-generation		0.211*** (0.042)	0.211*** (0.042)	0.183*** (0.042)	0.211*** (0.042)	0.199*** (0.055)
Second-generation		0.038+ (0.020)	0.038+ (0.020)	0.028* (0.013)	0.038+ (0.020)	0.023 (0.039)
SES		0.001 (0.011)	-0.034 (0.021)	0.009 (0.022)	-0.035+ (0.020)	-0.021+ (0.012)
Central exams		0.009 (0.139)	0.009 (0.139)	0.061 (0.129)	0.004 (0.127)	
Central exams × SES			0.051* (0.026)	0.056* (0.026)	0.052* (0.025)	0.038+ (0.020)
Performance (PVREAD)				-0.195*** (0.041)		
Tracking age					0.075** (0.023)	
Tracking age × SES					-0.009* (0.004)	
Country FE						
Constant	No	No	No	No	No	Yes
	0.803*** (0.056)	1.600** (0.568)	1.600** (0.568)	2.081*** (0.568)	1.604** (0.563)	1.289** (0.463)
Random part						
$\hat{\sigma}_{u0}^2$ (country-level)	0.168*** (0.044)	0.166*** (0.037)	0.166*** (0.037)	0.153*** (0.031)	0.148*** (0.031)	
$\hat{\sigma}_{v0}^2$ (school-level)	0.126*** (0.032)	0.110*** (0.027)	0.110*** (0.027)	0.109*** (0.029)	0.110*** (0.027)	0.116*** (0.013)
$\hat{\sigma}_e^2$ (student-level)	2.236*** (0.140)	2.234*** (0.140)	2.234*** (0.140)	2.216*** (0.141)	2.234*** (0.140)	2.236*** (0.041)
$\hat{\sigma}_{u6}^2$ (SES)		0.007*** (0.002)	0.006*** (0.001)	0.007*** (0.001)	0.006*** (0.001)	
BIC	13,980,521.9	13,964,970.8	13,964,979.0		13,964,995.6	13,972,982.3
Log likelihood	-6,990,236.0	-6,982,410.6	-6,982,408.5		-6,982,404.3	-6,986,098.3

Source: PISA 2012 own calculations.

SES: socio-economic status, FE: fixed effects, BIC: Bayesian information criterion.

Standard errors in parentheses. Normalized sampling weights are used (robust standard errors reported).

Tracking age is mean-centered. $N = 260,267$ (15,973 schools, 54 countries).

+ $p < 0.10$; * $p < 0.05$; ** $p < 0.05$; *** $p < 0.001$.

Table A3. Results multilevel regression models personal tutor (use intensity: hours a week).

	(0)	(1)	(2)	(3)	(4)	(5)
Fixed part						
Female		0.179*** (0.048)	0.179*** (0.048)	0.154** (0.051)	0.179*** (0.048)	0.180*** (0.026)
Age of student		0.111* (0.045)	0.111* (0.045)	0.099* (0.046)	0.111* (0.045)	0.110* (0.048)
Grade		-0.141*** (0.027)	-0.141*** (0.027)	-0.066** (0.024)	-0.141*** (0.027)	-0.142*** (0.026)
Immigrant background (ref. = native)						
First-generation		0.258*** (0.068)	0.258*** (0.068)	0.234*** (0.060)	0.258*** (0.068)	0.269*** (0.082)
Second-generation		0.133*** (0.024)	0.133*** (0.024)	0.123*** (0.029)	0.133*** (0.024)	0.141* (0.059)
SES		0.150*** (0.021)	0.110*** (0.033)	0.159*** (0.034)	0.109** (0.033)	0.058*** (0.014)
Central exams		0.091 (0.172)	0.091 (0.172)	0.157 (0.157)	0.090 (0.171)	
Central exams × SES			0.060 (0.044)	0.067 (0.044)	0.061 (0.044)	0.220*** (0.028)
Performance (PVMATH)				-0.214*** (0.044)		
Tracking age					0.013 (0.039)	
Tracking age × SES					-0.010 (0.009)	
Country FE	No	No	No	No	No	Yes
Constant	0.830*** (0.071)	0.322 (0.831)	0.322 (0.831)	0.804 (0.849)	0.323 (0.830)	0.890 (0.719)
Random part						
$\hat{\sigma}_{u0}^2$ (country-level)	0.265*** (0.053)	0.363*** (0.077)	0.363*** (0.077)	0.326*** (0.069)	0.362*** (0.077)	
$\hat{\sigma}_{v0}^2$ (school-level)	0.185*** (0.058)	0.147*** (0.044)	0.147*** (0.044)	0.149*** (0.046)	0.147*** (0.044)	0.159*** (0.016)
$\hat{\sigma}_e^2$ (student-level)	3.512*** (0.572)	3.466*** (0.562)	3.466*** (0.562)	3.441*** (0.567)	3.466*** (0.562)	3.477*** (0.083)
$\hat{\sigma}_{u6}^2$ (SES)		0.023*** (0.006)	0.022*** (0.006)	0.022*** (0.006)	0.022*** (0.006)	
BIC	13,727,427.5	13,665,090.1	13,665,100.8		13,665,124.7	13,682,118.0
Log likelihood	-6,863,689.0	-6,832,470.7	-6,832,469.9		-6,832,469.4	-6,840,668.7

Source: PISA 2012 own calculations.

SES: socio-economic status, FE: fixed effects, BIC: Bayesian information criterion.

Standard errors in parentheses. Normalized sampling weights are used (robust standard errors reported).

Tracking age is mean-centered. N=240,729 (15,914 schools, 54 countries).

*p < 0.10; **p < 0.05; ***p < 0.001.

Table A4. Results multilevel regression models commercial company lesson (use intensity: hours a week).

	(0)	(1)	(2)	(3)	(4)	(5)
Fixed part						
Female		0.128*** (0.027)	0.128*** (0.027)	0.114*** (0.028)	0.128*** (0.027)	0.128*** (0.029)
Age of student		0.159* (0.063)	0.159* (0.063)	0.153* (0.062)	0.159* (0.063)	0.157** (0.054)
Grade		-0.158*** (0.041)	-0.158*** (0.041)	-0.115** (0.040)	-0.158*** (0.041)	-0.155*** (0.028)
Immigrant background (ref. = native)						
First-generation		0.221** (0.081)	0.221** (0.081)	0.207** (0.074)	0.221** (0.081)	0.244** (0.089)
Second-generation		0.205*** (0.039)	0.205*** (0.039)	0.199*** (0.035)	0.205*** (0.039)	0.237*** (0.062)
SES		0.180*** (0.027)	0.127*** (0.034)	0.153*** (0.035)	0.127*** (0.034)	0.123*** (0.015)
Central exams		0.243 (0.270)	0.243 (0.270)	0.278 (0.259)	0.240 (0.268)	
Central exams × SES			0.078 (0.051)	0.081 (0.050)	0.078 (0.051)	0.130*** (0.032)
Performance (PVMATH)				-0.121* (0.048)		
Tracking age					0.035 (0.056)	
Tracking age × SES					-0.000 (0.011)	
Country FE	No	No	No	No	No	Yes
Constant	0.854*** (0.110)	-0.288 (0.747)	-0.288 (0.747)	-0.026 (0.714)	-0.287 (0.745)	0.143 (0.789)
Random part						
$\hat{\sigma}_{u0}^2$ (country-level)	0.643 (0.201)	1.001 (0.371)	1.001 (0.371)	0.979 (0.373)	0.997 (0.367)	
$\hat{\sigma}_{v0}^2$ (school-level)	0.317*** (0.067)	0.267*** (0.055)	0.267*** (0.055)	0.272*** (0.057)	0.267*** (0.055)	0.293*** (0.020)
$\hat{\sigma}_e^2$ (student-level)	4.168*** (0.620)	4.126*** (0.614)	4.126*** (0.614)	4.117*** (0.616)	4.126*** (0.614)	4.137*** (0.094)
$\hat{\sigma}_{u6}^2$ (SES)		0.038*** (0.010)	0.037*** (0.010)	0.036*** (0.009)	0.037*** (0.010)	
BIC	14024181.6	13974479.8	13974490.4		13974514.9	13992018.9
Log likelihood	-7012066.0	-6987165.6	-6987164.7		-6987164.6	-6995619.5

Source: PISA 2012 own calculations.

SES: socio-economic status, FE: fixed effects, BIC: Bayesian information criterion.

Standard errors in parentheses. Normalized sampling weights are used (robust standard errors reported). Tracking age is mean-centered. $N=237,737$ (15,899 schools, 54 countries).

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.001$.